Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Site-specific diel patterns and short-term predictability in fishway video annotations

Miracle Amadi

miracle.amadi@lut.fi

Lappeenranta University of Technology

David Nijssen

Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, BfG)

Cornelia Schütz

Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, BfG)

Jeffrey A. Tuhtan

Tallinn University of Technology

Article

Keywords: fishway monitoring, fish passage, underwater video, diel activity, Gaussian process, time series forecasting, annotation quality control

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9681371/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Site-specific diel patterns and short-term predictability in fishway video annotations

Miracle Amadi[1], David Nijssen[2], Cornelia Schütz[2], and Jeffrey A. Tuhtan[3]

[1]LUT School of Engineering Sciences, Lappeenranta–Lahti University of Technology (LUT), Yliopistonkatu 34, FI-53850 Lappeenranta, Finland, miracle.amadi@lut.fi
[2]Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, BfG), Koblenz, Germany, nijssen@bafg.de, schuetz@bafg.de
[3]Centre for Environmental Intelligence and Sensing, Department of Computer Systems, Tallinn University of Technology, Akadeemia tee 15A, 19086 Tallinn, Estonia, jeffrey.tuhtan@taltech.ee

Abstract

Fixed underwater cameras are increasingly used to monitor fish migration, but the resulting video archives are large, heterogeneous, and costly to annotate. Human video annotations remain the most reliable ground truth for fish species, size, and behaviour, and can also provide site-specific reference patterns for validating human and machine learning-based outputs. Here, we analyse the potential and limitations of human annotations as availability-aware time series, where only hours with available video observations are analysed, using data from three fishway monitoring sites in large freshwater rivers in Germany. We examined size-behaviour and species-behaviour relationships, diel hour-of-day activity patterns, and the consistency of these patterns across sites with different species composition, abundance, and recording continuity. To assess short-term predictability, we constructed hourly count series and fitted Gaussian process models using only past observations to forecast 24-hour fish passage. Across all sites, fish passage showed a strong and consistent diel pattern, with activity concentrated between approximately 04:00 and 18:00, including peaks around 04:00–06:00 and 14:00–17:00, and was dominated by Small Cyprinids and upstream movements. Size and species effects primarily influenced the magnitude of observed activity rather than its timing, and similar temporal structures were observed despite otherwise large site-to-site differences. Forecasts captured broad diel structure but had limited ability to localise brief, spike-like passage events when recent observations were sparse. Predictive uncertainty was generally well calibrated, and performance depended more on recording continuity than on the total length of the training history. These results establish a transparent statistical baseline for hourly forecasting and show how interpretable site-specific patterns can support annotation quality control.

keywords: fishway monitoring, fish passage, underwater video, diel activity, Gaussian process, time series forecasting, annotation quality control

1 Introduction

Climate change, together with hydropower development and other human activities, is transforming aquatic ecosystems at an alarming pace [1, 2]. In freshwater systems, altered flow regimes, rising temperatures, and river fragmentation disrupt the timing and success of fish migration, particularly for species that depend on short seasonal windows for spawning and

passage [3–5]. These pressures interact locally with dams and other infrastructure, making it essential to monitor when fish pass, which species are present, and how passage intensity varies at fine temporal scales [6].

Rivers and coastal systems are increasingly monitored with fixed underwater cameras and other optical sensors [7, 8]. In practice, these systems generate vast archives of raw video that still need to be interpreted by humans or machine learning (ML) models [9, 10]. The accuracy of taxonomic classification in these recordings can be enhanced by incorporating spatiotemporal context, such as known migration patterns, expected species assemblages, and size distributions, particularly when these data are available at hourly or finer temporal resolution [11]. Yet analysing reliable time series at this scale remains a challenge. Fish passage is rarely homogeneous in time; instead, movement patterns often occur in short, clustered bursts [12], so migration intensity can shift markedly within a few hours [13]. Capturing the variability in species and size composition as well as in the resulting time series, is essential to evaluate the reliability of background information used to enhance taxonomic classification [14].

At present, the most reliable information comes from expert human annotators [15] who label individual fish in videos by species, size, and behaviour. Manual taxonomic annotation of video footage remains an exhaustive process requiring substantial labour and specialised expertise; it is particularly challenging in riverine environments where visibility is often reduced by turbidity and variable illumination, rendering human error unavoidable. Nevertheless, they form a critical "ground truth" against which automated systems are designed and evaluated. A method to support manual annotation through site-specific analysis of annotated data to identify and reduce errors is still missing. While annotations are used to label individual fish, they also contain structured information on species occurrence, size distributions, and movement behaviour that can be analysed collectively [16]. Such methods should be based on: (i) statistical analysis of behaviour and occurrence of species in the human ground-truth annotations within a specific site, (ii) comparison of the results across sites, and (iii) evaluation of the reliability for hourly forecasting of fish passage.

At the same time, automated ML systems are increasingly deployed to analyse these video streams. For example, Shafait et al. [17] demonstrated automated fish species recognition in unconstrained underwater environments using image-set classification, achieving high accuracy without manual feature extraction. Recently, Owen [18] developed a real-time monitoring system based on convolutional neural networks that enables species recognition and behavioural analysis directly from underwater video streams. However, without a clear statistical benchmark for how well such data can be predicted from reference dataset alone, it is difficult to know whether failures arise from the ML model, the underlying variability, or gaps in the observational record.

Recent freshwater fish datasets developed for computer vision vary widely in scale, taxonomic breadth, and recording context. For example, a North American salmon-monitoring dataset contains approximately 532,000 labelled frames from weirs in British Columbia, Canada [19]. The Angling Freshwater Fish Netherlands (AFFiNe) dataset includes around 7,000 angling photographs of 30 common freshwater species [20], collected from angling photographs (ex situ), while the PTFish dataset provides about 18,500 annotated frames for fish detection and length estimation from two natural river sites in Portugal [21]. The KakaduFishAI dataset contributes roughly 44,000 underwater images covering 23 tropical freshwater species from river and floodplain habitats in Kakadu National Park, northern Australia [22]. These datasets have been developed primarily to train and evaluate machine learning models for tasks such as detection, tracking, or species classification. Additional summaries of freshwater fish detection datasets have been reported in recent unpublished work [23]. The present study uses human annotations from more than 400,000 short underwater videos (typical duration < 3 min.) recorded at the exit of three fishways in Germany by the German Federal Institute of Hydrology (BfG). Each fish observation is time-stamped and annotated with species identity, estimated body size, and migration behaviour, allowing fish passage to be analysed as a time series of observations. To our knowledge, this represents one of the largest European freshwater fish annotation datasets reported.

The overall aim of this work is to provide a methodology that can be used to improve taxonomic identification by human and/or ML annotators. Our first aim is to characterise the structure and patterns of these annotations within each site, as a first step toward hourly prediction of fish

counts. We examine how body size relates to behaviour (direction of movement) within species, how species and behaviour combine to produce site-specific patterns of presence and movement, and how diel (hour-of-day) activity differs between size classes and species. These site- and time-specific factors are central to supporting manual and improving ML annotation for these sites, since body size strongly influences swimming capacity and passage performance [12, 24], while movement direction reflects how fish enter and leave the camera field [25]. Diel activity patterns further structure when fish are observed, often producing recurring temporal patterns within a 24-hour cycle [12]. This provides a data-driven description of site-specific migration dynamics in the video sequences [26]. For example, it allows us to identify which combinations of species, size, and behaviour are common, rare, or effectively absent in the videos, offering a practical way to flag implausible or inconsistent annotations and/or model predictions [27, 28].

Our second aim is to develop a model to predict hourly fish counts based on the data characterisation in the first step, with a focus on site-specific short-term history. To achieve this, we use a Gaussian process regression model in which time is the only predictor to forecast counts over the next 24 hours for selected species and size groups at each site. By evaluating these forecasts across multiple time windows, we quantify how prediction accuracy and uncertainty vary through time within each site. This provides a transparent method that can be applied to other sites to support and optimise taxonomic identification by human annotators or ML annotations.

To conclude our analysis and explore parallels between all three sites, we analyse fish-level annotations from three monitoring sites that differ in species composition, monitoring duration, and the proportion of hours for which videos were available. Each annotation describes a single observed fish with species names, estimated body size, and one of six defined directional behaviours indicating how the fish entered and exited the camera field. From these data, we construct availability-aware hourly count series, using only hours with available video observations, for common species and size groups, and analyse the resulting time series to characterise the statistical structure present within each site.

2 Materials and methods

This section describes the human fish-level annotations used in this study and the steps taken to prepare them for analysis. Each annotation corresponds to an observed fish occurrence and is represented by a set of categorical and continuous attributes, including species label, estimated body size, and directional swimming behaviour.

2.1 Description of sites

This study is based on video monitoring data from three fishways located in freshwater river systems: Rothenfels on the River Main (Site 1), Malliß on the Müritz-Elde-Waterway (Site 2), and Koblenz on the River Mosel (Site 3). The datasets differ in monitoring duration, intensity of recording, and species composition. Across all three sites, fixed underwater cameras recorded the passage of fish through defined viewing areas. Videos were recorded as short clips of up to three minutes in duration. The videos were collected by the German Federal Institute of Hydrology (BfG) and processed by the Environmental Sensing and Intelligence team at Tallinn University of Technology (TalTech). Table 1 summarises the monitoring periods, number of videos, annotated fish entries, and observed hours used in the analyses.

After preprocessing, the annotated datasets comprise 4,350 to 31,833 videos per site and between 8,740 and 172,813 individual fish observations. A fish entry corresponds to a single annotated observation of an individual fish within a video clip. The number of observed hours available for analysis ranges from 1,023 to 4,456 hours per site.

The most frequently observed species differ across sites but are primarily Cyprinids, including Roach (*Rutilus rutilus*), Chub (*Squalius cephalus*), Bleak (*Alburnus alburnus*), and Small Cyprinids (group-level classification), with additional site-specific dominance by European nase (*Chondrostoma nasus*) or Barbel (*Barbus barbus*), and by the percid European perch (*Perca fluviatilis*).

Table 1: Summary of annotated data after preprocessing. Dominant species are defined as the five species with the highest total number of annotated individuals at each site.

Site	Monitoring period	No. of videos	No. of fish entries	Observed hours	Top 5 species
Site 1	2020-06-23 to 2021-06-30	31,233	135,949	3,157	Roach, Chub, European nase, Bleak, European perch
Site 2	2019-06-03 to 2019-11-27	4,350	8,740	1,048	Roach, European perch, Bleak, Small Cyprinids, Chub
Site 3	2020-04-01 to 2023-09-18	31,833	172,813	4,456	Small Cyprinids, Bleak, Barbel, Roach, European perch

2.2 Fish-level annotation components

Fish-level annotations contain the following components:

- **Species**: Human-identified fish species (e.g., Roach, Chub, European eel, Barbel). Some labels reflect uncertainty or coarse group-level classification; for example, *Unable to clearly identify* denotes unresolved identification (Fig. 5b), while *Small Cyprinids* is not a species label but a grouping used for Cyprinids that are too small (typically below approximately 10–15 cm) to be reliably identified to species level.
- **Size**: Human-estimated fish size recorded in approximate 5 cm increments.
- **Behaviour**: One or more of six predefined directional behaviours indicating swimming pattern and entry/exit: `UP`, `DOWN`, `IN_UP`, `IN_DOWN`, `UP_IN`, and `DOWN_IN`. Behaviour is defined relative to the camera field of view and the direction of movement through the image frame, and may also reflect differences in how fish move within the water column. Behaviour definitions are provided in Supplementary Material 1.

2.3 Preprocessing, timestamps, and availability-aware counts

For each site, we restricted the dataset to videos in which at least one fish was annotated with one of the six defined directional behaviours. Videos that contained none of these behavioural events were excluded. Such cases may include fish that are present but not moving directionally (e.g. waiting, schooling, hovering, feeding, or reacting to light). These were excluded because the focus is on directional movement through the fishway, where the primary interest is the net movement of fish rather than the time spent within the camera field of view without directional movement.

Some annotation rows report more than one individual (i.e., the count field is greater than 1). These rows typically correspond to cases where the annotator observed multiple fish of the same species, size, and behaviour within the same video. To obtain a consistent fish-level dataset, we expanded such rows so that each row corresponds to exactly one fish. After expansion, we removed records with undefined or invalid behaviour labels. The final fish-level dataset therefore contains one row per observed fish instance.

Each fish annotation includes start and end times within the source video, but only one of these is used to define the timing of the behaviour. During annotation, raters manually adjust either the start or the end time depending on behaviour type, while the other field remains at the original video boundary. For exit-type behaviours (`UP`, `DOWN`, `IN_UP`, `IN_DOWN`), the end time marks the behaviour, whereas for entry-type behaviours (`UP_IN`, `DOWN_IN`), the start time is used. Because original video boundaries are not stored explicitly, behaviour timing is inferred solely from the edited (manually adjusted) timestamp. To construct hourly time series, we combine the calendar date inferred from the video file metadata with the hour-of-day derived from this behaviour time. The resulting timestamps are discretised to hourly resolution by flooring to the nearest hour, so

that each fish is assigned to a unique hourly bin.

Video recording was not continuous at any site, as cameras were motion-activated to provide a preliminary filter of empty videos without fish activity. In addition, recording was occasionally halted (e.g. during periods of low visibility, particularly at Sites 1 and 2). As a result, although monitoring periods span months to years, there are hours or even full days for which no video data is available. All hourly analyzes are therefore conducted using the set of hours for which video data exist and fish observations are available. This availability-aware handling of missing data conditions forecasts on observed hours rather than imputing values for periods without recordings, as is commonly done in ecological time-series analysis [29, 30].

2.4 Annotation structure analysis

We used the human annotations to quantify how body size, movement behaviour, species identity, and time-of-day are associated with observed fish passage at each site. All analyses were carried out separately by site to avoid conflating between-site differences with variation in monitoring coverage within sites.

Size–behaviour relationships. For each site and each of the six movement behaviours, we summarised body size (cm) using sample size (N), mean, standard deviation, quartiles, and range (Table 2). Distributions were visualised using violin plots with consistent axis scaling within each site. In addition to the six annotated behaviour categories, behaviours were also aggregated into upstream-oriented (`UP`, `IN_UP`, `UP_IN`) and downstream-oriented (`DOWN`, `IN_DOWN`, `DOWN_IN`) movement groups to compare broader directional structure across sites.

Species–behaviour relationships. For each site, we constructed species–behaviour contingency tables containing the total number of annotated individuals for each species–behaviour combination. These were visualised as heatmaps. Species were ordered by total abundance within each site to highlight frequently observed species and rare combinations.

Because raw counts are strongly influenced by species abundance, we additionally computed species-normalised behaviour proportions (row-wise percentages). These heatmaps show the relative distribution of behaviours within each species, allowing direct comparison of behavioural tendencies across species with different sample sizes.

Diel (hour-of-day) activity patterns. Diel structure was analysed using availability-aware hourly counts (Section 2.3). For each site, hourly totals were computed for (i) all fish combined, (ii) behaviour, (iii) size class (fish lengths were grouped into three fixed bins applied consistently across sites: *Small* (≤ 15 cm), *Medium* (16–30 cm), and *Large* (≥ 31 cm)), and (iv) species.

To summarise diel trends, we fitted generalised additive models (GAMs) to hourly count data using hour of day as the only predictor [31, 32]. Hour of day was used as a consistent temporal index across the full monitoring period. Accordingly, the fitted patterns describe variation by clock time rather than by ecologically defined light phases. Diel activity patterns are expected to be smooth but not necessarily symmetric or unimodal, and may exhibit multiple peaks within a 24-hour cycle. Parametric approaches such as harmonic regression impose a fixed sinusoidal form, which can be too restrictive for such patterns. GAMs instead estimate a smooth function of hour of day directly from the data, allowing flexible representation of complex diel structure while retaining interpretability [32, 33].

In this framework, hourly counts y_t are treated as conditionally independent Poisson observations with mean λ_t, linked to hour of day via a smooth function:

$$y_t \sim \text{Poisson}(\lambda_t), \qquad \log(\lambda_t) = f(h_t),$$

where h_t denotes hour of day and $f(\cdot)$ is represented using spline basis functions with an

associated smoothing penalty [32]. The degree of smoothness is estimated from the data, providing a balance between fidelity to observed counts and avoidance of overfitting.

Models were fitted in Python using `statsmodels` (`GLMGam` with `BSplines`), with a Poisson error distribution and log link function. Because counts were aggregated across the full monitoring period, the fitted smooth represents the expected count at each hour of day and can be interpreted as a smooth diel intensity function describing how fish passage varies across the 24-hour cycle. Points show pooled counts per hour of the day and lines show the GAM fitted smooth with a confidence interval of 95%, derived from the standard errors of the fitted model.

To quantify goodness-of-fit between observed hourly counts and the GAM-smoothed diel curves, we additionally computed the root mean squared error (RMSE) between observed and fitted hourly counts:

$$\mathrm{RMSE} = \sqrt{\frac{1}{24}\sum_{h=1}^{24}(y_h - \hat{y}_h)^2},$$

where y_h denotes the observed pooled count at hour h and $\hat{y}_h$ the corresponding GAM fitted value. Because overall count magnitudes differed substantially among sites, we also report the normalised RMSE (NRMSE),

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\bar{y}},$$

where $\bar{y}$ is the mean observed hourly count for that site. Normalisation allows goodness-of-fit to be compared across sites with different overall activity levels.

2.5 Forecasting model

We assess how predictable hourly fish passage counts are when we use the past history of human annotations. We present a forecasting structure that can be applied consistently across sites and targets. We focus on three illustrative target types that occur consistently across sites and represent different levels of aggregation: (i) a common species (*Roach*), (ii) a dominant size class (`Small`), and (iii) a species–size subgroup (*Roach*×`Small`). These examples are used to demonstrate typical model behaviour under well-supported, size-aggregated, and more specific but sparser conditions, respectively; results for additional targets can be generated using the same procedure.

Observed-hours and hourly count targets. Video recordings were not continuous at any site, so hourly counts were modelled using only hours with available video observations, rather than inserting zeros for missing periods. For each site, we define the set of observed hours as

$$\mathcal{T}_{\mathrm{obs}} = \{t : \text{at least one annotation exists in hour } t\},$$

and construct a target-specific hourly count series

$$y_t^{(g)} = \#\{\text{fish annotated with target } g \text{ in hour } t\}, \qquad t \in \mathcal{T}_{\mathrm{obs}}.$$

If a target is absent during an observed hour, $y_t^{(g)} = 0$. Hours with no video or annotations are treated as missing and excluded. This avoids conflating camera downtime with true absence of fish and ensures that forecasts are evaluated only when observations were possible.

Gaussian process model for hourly counts. For each target g, we model log-transformed hourly counts,

$$z_t^{(g)} = \log\bigl(1 + y_t^{(g)}\bigr),$$

using Gaussian process (GP) regression [34, 35], with time (in hours since the start of the dataset) as the only input. Thus, the dependent variable is the log-transformed hourly count $z_t^{(g)}$, and the independent variable is time t.

GPs were employed because they provide a flexible, non-parametric framework for modelling smooth temporal processes without specifying a fixed functional form, while naturally quantifying predictive uncertainty. This is appropriate here because hourly fish passage exhibits strong temporal variation with both regular components (e.g., diel periodicity) and irregular fluctuations, but no simple parametric model is available. The Gaussian process defines a prior over functions $f(t)$ through the choice of covariance kernel, allowing this variation to be represented as a combination of periodic (diel), slowly varying, and short-term components.

The latent function is modelled as

$$f(t) \sim \mathcal{GP}(0, k(t, t')),$$

with observations given by

$$z_t = f(t) + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2_{\text{noise}}).$$

The covariance kernel is defined as

$$k(t, t') = \sigma^2 \left(k_{\text{short}} + k_{\text{long}} + k_{\text{per}}\right),$$

where k_{per} is a periodic kernel with a 24-hour period, and k_{short} and k_{long} are squared exponential (radial basis function, RBF) kernels with different length scales representing short- and long-term temporal variation, respectively. The periodic component is motivated by the strong and consistent diel patterns observed across sites (Section 3.1.3). The kernel components have separate characteristic scales, while their overall contribution is controlled by a shared amplitude parameter σ^2. Residual variation is modelled through the Gaussian noise term ϵ_t. All hyperparameters were estimated from the data by maximising the marginal likelihood of the Gaussian process.

The GP assumes Gaussian residual variation on the log-transformed scale. Raw counts are not assumed to be Gaussian; instead, approximate normality is assumed after the $\log(1 + y)$ transformation. The addition of 1 ensures that zero counts can be included without undefined values.

For long training histories, we apply regular subsampling (thinning) of training inputs to maintain computational tractability while keeping the GP likelihood, kernel structure, and inference procedure unchanged [36, 37]. The thinning interval is chosen adaptively based on the size of the training set. Specifically, if N training points are available, we retain every s-th point, where $s = \lceil N/10^3 \rceil$, so that the effective training set remains on the order of 10^3 points.

Predictions are transformed back to the original count scale using the log-normal mean approximation [38, 39],

$$\widehat{y}_t = \exp\left(\mu_t + \tfrac{1}{2}\sigma_t^2\right) - 1,$$

and uncertainty is reported with transformed 95% predictive intervals. All Gaussian process models were implemented in Python (version 3.12.3) [40] using `scikit-learn` (version 1.5.2) [41].

Test window selection and evaluation. To provide a comparable example across sites (despite different monitoring periods and recording gaps), we present forecasts for a single 24-hour test window per site. The window was selected near the end of each dataset as the *densest* 24-hour period, defined as the near-end window with the highest observed-hour coverage that met a minimum coverage requirement. Each test window was required to contain at least 16 observed hours (approximately two-thirds of a full day). This threshold was chosen as a pragmatic minimum coverage rule to ensure that validation metrics were computed over a sufficiently large portion of the diel cycle and were not dominated by a few isolated observations. Lower thresholds would retain more windows but yield less stable estimates of MAE, correlation, total relative error, and interval coverage, whereas higher thresholds would exclude many otherwise usable windows, particularly at intermittently recorded sites. Training used all observed hours up to the cutoff time $t_{\text{cut}} = t_{\text{end}} - 24, \text{h}$, and the test set consisted of observed hours in $(t_{\text{cut}}, t_{\text{end}}]$.

Performance metrics. We summarise forecast performance using four complementary metrics computed on the test window: mean absolute error (MAE), Pearson correlation between observed and predicted counts, relative error in total counts over the window, and empirical coverage of the nominal 95% prediction interval. Definitions and aggregation across multiple test windows are given in Section 2.6.

2.6 Validation strategy

A single near-end 24-hour window is useful for visual comparison (as in Fig. 10, for example), but it can be unrepresentative if that day is unusually quiet or unusually active. We therefore validate the forecasting pipeline using multiple rolling 24-hour test windows near the end of each site's time series, following standard time-series cross-validation procedures [42].

Rolling windows are generated from the observed-hour index $\mathcal{T}_{\text{obs}}$ (hours with video/annotations). For each site, we start from the last observed hour in the dataset and step backwards in 24-hour increments. Each candidate window is defined by its end time t_{end} and a 24-hour cutoff

$$t_{\text{cut}} = t_{\text{end}} - 24\,\text{h}.$$

The test set for that window is the set of recorded hours within $(t_{\text{cut}}, t_{\text{end}}]$,

$$\mathcal{T}_{\text{test}} = \{t \in \mathcal{T}_{\text{obs}} : \; t_{\text{cut}} < t \leq t_{\text{end}}\}, \qquad n_{\text{test}} = |\mathcal{T}_{\text{test}}|.$$

Thus, each test-set data point corresponds to a single observed hour with available video data within the 24-hour evaluation window, together with the associated hourly count for the target being forecast.

A window is *retained* only if it contains sufficient recording coverage (here, at least 16 observed hours; i.e., $n_{\text{test}} \geq 16$). Otherwise it is skipped and the algorithm moves to the next candidate t_{end} one day earlier. We retain up to 30 accepted windows per site (or it can be fewer if insufficient near-end coverage is available).

For each accepted window, the model is trained on observed hours up to and including t_{cut} (i.e., the full available history prior to the test window). To ensure the GP is not fit on an extremely short history, we also require a minimum training size (here, at least 36 observed hours prior to the cutoff), but for all accepted rolling windows across sites, the effective training sets contain several hundred to several thousand observed hours. This guarantees that each evaluation window is based on a non-trivial training record rather than a very small fragment of data.

For each retained window we compute four complementary metrics and then summarise them across windows using the median and interquartile range. Let y_t be the observed count and $\hat{y}_t$ the GP predictive mean for hours $t \in \mathcal{T}_{\text{test}}$. We report:

$$\text{MAE} = \frac{1}{n_{\text{test}}} \sum_{t \in \mathcal{T}_{\text{test}}} |y_t - \hat{y}_t|,$$

the correlation between $\{y_t\}$ and $\{\hat{y}_t\}$ over $t \in \mathcal{T}_{\text{test}}$ (a timing/shape diagnostic), the total relative error in the 24-hour sum,

$$\text{TRE} = \frac{\sum_{t \in \mathcal{T}_{\text{test}}} \hat{y}_t - \sum_{t \in \mathcal{T}_{\text{test}}} y_t}{\sum_{t \in \mathcal{T}_{\text{test}}} y_t},$$

and the empirical coverage of the nominal 95% predictive interval,

$$\text{Coverage}_{95} = \frac{1}{n_{\text{test}}} \sum_{t \in \mathcal{T}_{\text{test}}} \mathbf{1}\{y_t \in [\ell_t, u_t]\},$$

where ℓ_t and u_t are the lower and upper bounds of the predictive interval at hour t. TRE is informative as a measure of bias in the 24-hour total, with positive values indicating overestimation and negative values indicating underestimation, but it can become unstable when the observed

total $\sum y_t$ is small; we therefore interpret it alongside MAE and coverage.

3 Results

3.1 Annotation structure across sites

3.1.1 Size–behaviour relationships across sites

Fish body sizes varied with migration behaviour across all three monitoring sites (Table 2; Fig. 1). While absolute size ranges differed between sites, similar size–behaviour patterns were observed across sites despite differences in species composition. When behaviours were aggregated into upstream- and downstream-oriented movement groups (Fig. 2), both directions were dominated by smaller fish across all sites. Differences between movement directions were expressed mainly in the spread of body sizes and the relative occurrence of larger individuals rather than in strong shifts in the most common body size.

Table 2: Descriptive statistics of fish size by migration behaviour, shown separately for each site. Fish lengths were estimated by annotators in approximate 5 cm increments. All size statistics are given in cm.

(a) Site 1

Behaviour	N	Mean	SD	Min	25%	50%	75%	Max
DOWN	25,264	29.73	13.30	5	20	30	40	100
DOWN_IN	9,559	27.28	13.16	5	20	25	40	65
IN_DOWN	14,356	26.41	12.83	5	15	25	40	100
IN_UP	15,638	24.88	12.68	5	15	20	30	100
UP	49,611	23.38	12.96	5	15	20	35	100
UP_IN	21,601	24.99	12.75	5	15	25	35	100

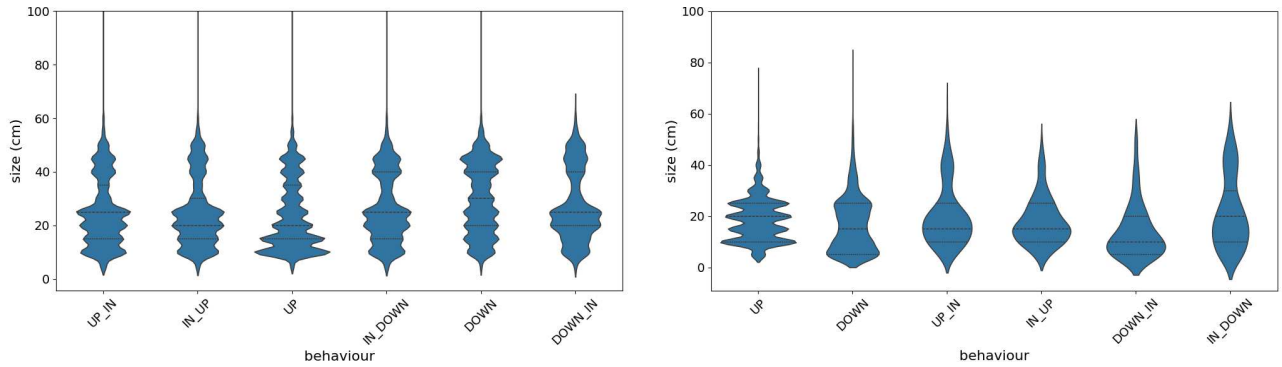
(b) Site 2

Behaviour	N	Mean	SD	Min	25%	50%	75%	Max
DOWN	1,626	17	10.82	5	5	15	25	80
DOWN_IN	118	14	10.24	5	5	10	20	50
IN_DOWN	207	22	13.86	5	10	20	30	55
IN_UP	250	18	9.18	5	10	15	25	50
UP	6,157	19	8.24	5	10	20	25	75
UP_IN	384	20	11.56	5	10	15	25	65

(c) Site 3

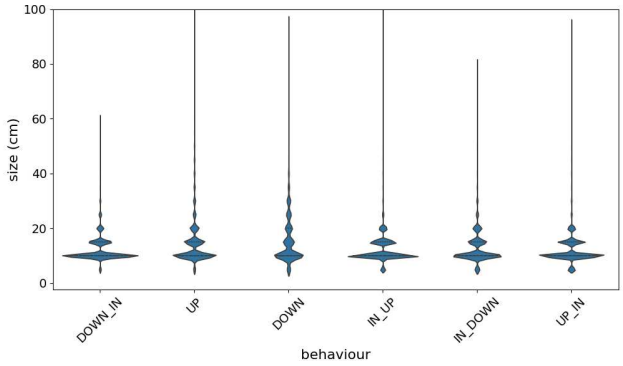
Behaviour	N	Mean	SD	Min	25%	50%	75%	Max
DOWN	15,543	15.25	8.17	5	10	10	20	95
DOWN_IN	11,862	12.19	4.26	5	10	10	15	60
IN_DOWN	12,734	12.77	5.49	5	10	10	15	80
IN_UP	33,598	12.29	4.66	5	10	10	15	100
UP	66,541	14.52	7.92	5	10	10	15	100
UP_IN	32,541	12.03	4.85	5	10	10	15	95

Across all sites, migration behaviours were dominated by smaller individuals, as indicated by the strong density peaks at lower sizes in the violin plots. Smaller fish were observed across all types of migration behaviour, including both upstream and downstream movements. Several distributions exhibited multiple local density peaks, most clearly at Sites 1 and 3 and to a lesser extent at Site 2, indicating multimodal size distributions within some behaviour categories. Differences between behaviours are not clearly expressed as shifts in the most common body size, but rather in the spread of sizes and the relative occurrence of larger individuals.



(a) Site 1

(b) Site 2

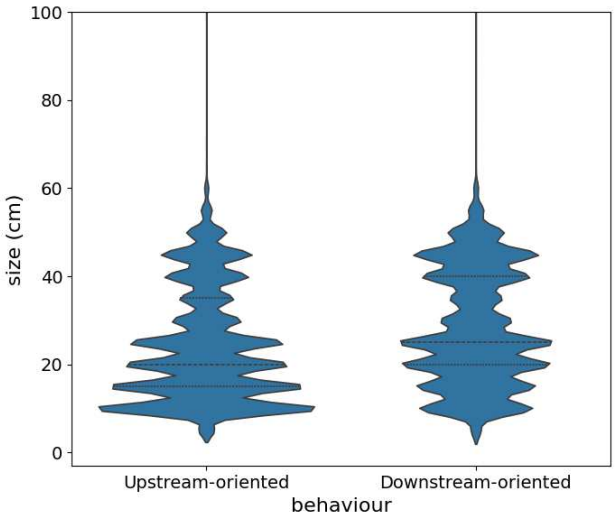


(c) Site 3

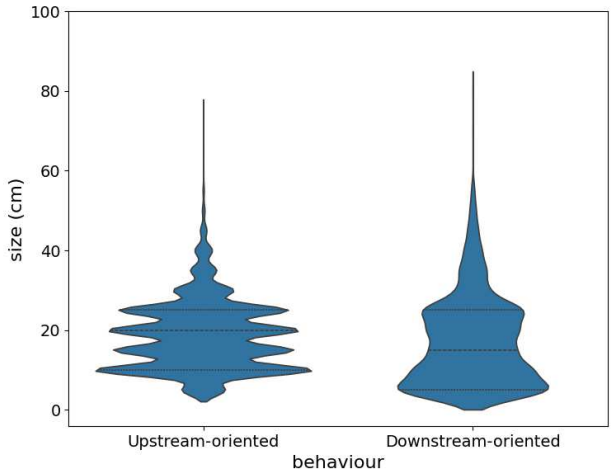
Fig. 1: Violin plots showing fish size distributions (cm) by migration behaviour for each site. Wider sections indicate a higher density of observations at a given size. Dashed lines represent the median and interquartile range (25% and 75% quantiles).

Differences between behaviours are most evident in the median and upper parts of the distributions (Table 2). Behaviours involving downstream movement out of the camera view (`DOWN`, `IN_DOWN`) and partial upstream entry (`UP_IN`) tended to include larger individuals more frequently than downstream entry-only behaviours (`DOWN_IN`). Complete upstream passages (`UP`) generally spanned a wide range of body sizes, and often included some of the largest individuals observed at a site.

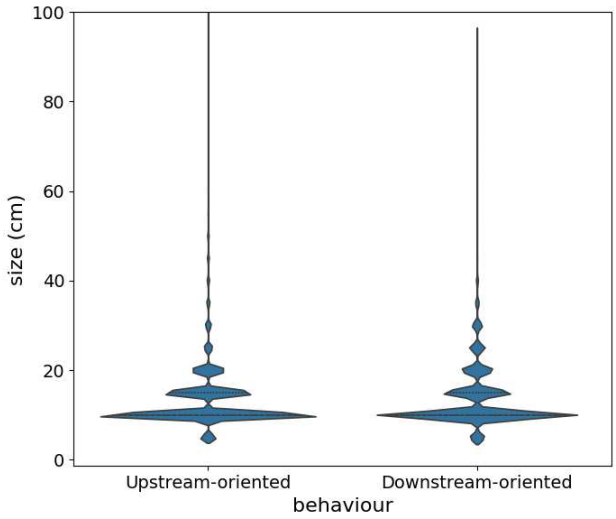
The strength and spread of these size–behaviour patterns varied by site. Site 1 showed broad size distributions within most behaviours, indicating a wide range of body sizes represented in the observations. In contrast, Site 3 exhibited much narrower distributions across behaviours, with medians concentrated at smaller sizes, reflecting a community dominated by small fish. Despite these differences in absolute scale, the relative ordering of behaviours by typical body size was broadly similar across sites.



(a) Site 1



(b) Site 2



(c) Site 3

Fig. 2: Aggregated fish size distributions (cm) for upstream-oriented (`UP`, `IN_UP`, `UP_IN`) and downstream-oriented (`DOWN`, `IN_DOWN`, `DOWN_IN`) movements across sites. Wider sections indicate higher observation density. Dashed lines represent the median and interquartile range (25% and 75% quantiles).

Overall, the results show that small fish are common in all behaviours, but behaviours differ in how widely sizes are distributed and in the presence of larger individuals. This consistent structure provides a useful reference for later analyses and for checking automated annotations, where unlikely size–behaviour combinations can point to possible errors. Formal comparisons of fish size among behaviour categories in the pooled dataset indicated statistically detectable differences in size distributions (one-way ANOVA and Kruskal–Wallis tests, all $p < 0.001$), although the distributions showed substantial overlap across behaviours.

3.1.2 Species–behaviour relationships across sites

The relationship between fish species and migration behaviour was examined using species–behaviour heatmaps for each site (Figure 3).

Across all sites, Cyprinids species such as Roach, Bleak, Chub, and Small Cyprinids contribute the majority of observations and occur across all migration behaviours. Despite differences in species

composition and overall abundance between sites, consistent patterns emerge: complete upstream passage (`UP`) is the most frequent behaviour for the most abundant species. In contrast, downstream-oriented behaviours (`DOWN` and `IN_DOWN`) are more evenly distributed across species.

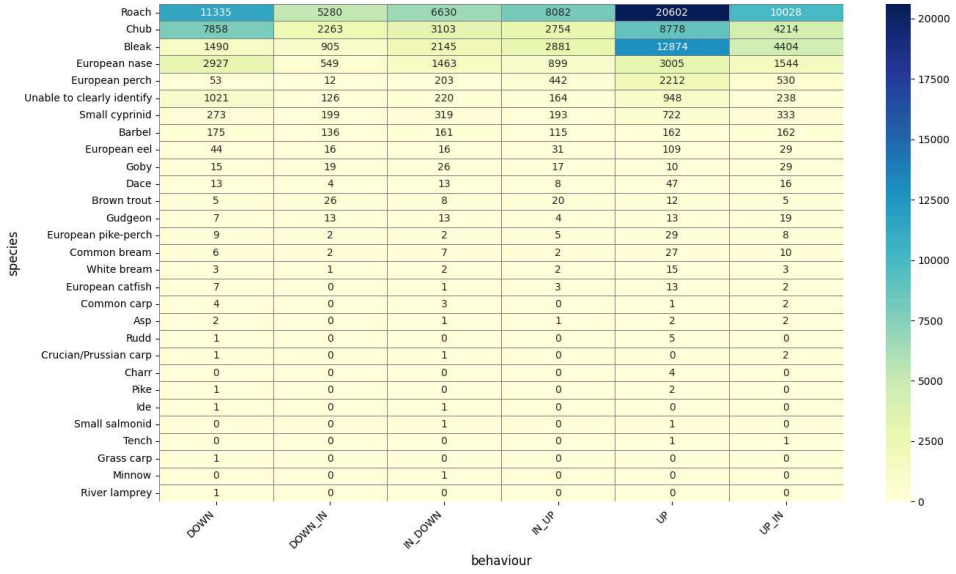
To account for differences in species abundance, species-normalised behaviour proportions were also examined (Figure 4). These reveal behavioural differences between species that are less apparent in the raw counts. For example, `DOWN` accounts for a larger share of observations in Chub than in Bleak across the sites, whereas Bleak is more strongly dominated by `UP`. Roach also shows a high proportion of `UP` movements, particularly at Sites 2 and 3. In contrast, Small Cyprinids show a more mixed pattern, especially at Site 3, where partial movements also contribute substantially.

Behaviours involving entry without exit (`DOWN_IN`, `UP_IN`) occur less frequently overall and are distributed across multiple species, with no single species strongly dominating these categories. This pattern suggests that these partial movements are not strongly species-specific, but reflect local movement within the camera field and appear mostly due to artefacts of the recording process (e.g. video clips ending before a full passage is observed).

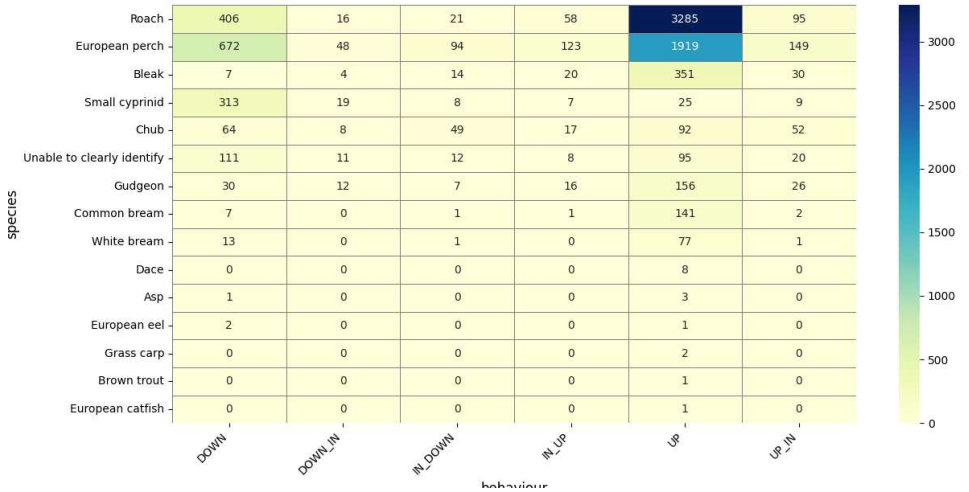
Entries labelled as *Unable to clearly identify* are present at all sites and across all behaviours, with higher counts typically associated with downstream movements. This likely reflects the difficulty of reliably identifying very small or fast-moving fish, which are more commonly observed during downstream passages; for example, at Site 2 these unclear identifications are concentrated at the smallest sizes (Figure 5b). Labels such as *Small Cyprinids* represent group-level identifications and are not included in Figure 5b.

Less frequently observed or large species such as European eel, European catfish, salmonids, and lampreys occur infrequently and are usually observed during complete upstream or downstream passages. Although their low counts limit statistical analysis, this pattern may reflect the relationship between body size and swimming performance. Larger fish generally swim faster, and are therefore more likely to cross the camera field of view before the end of a video (maximum duration 3 minutes) [43, 44].

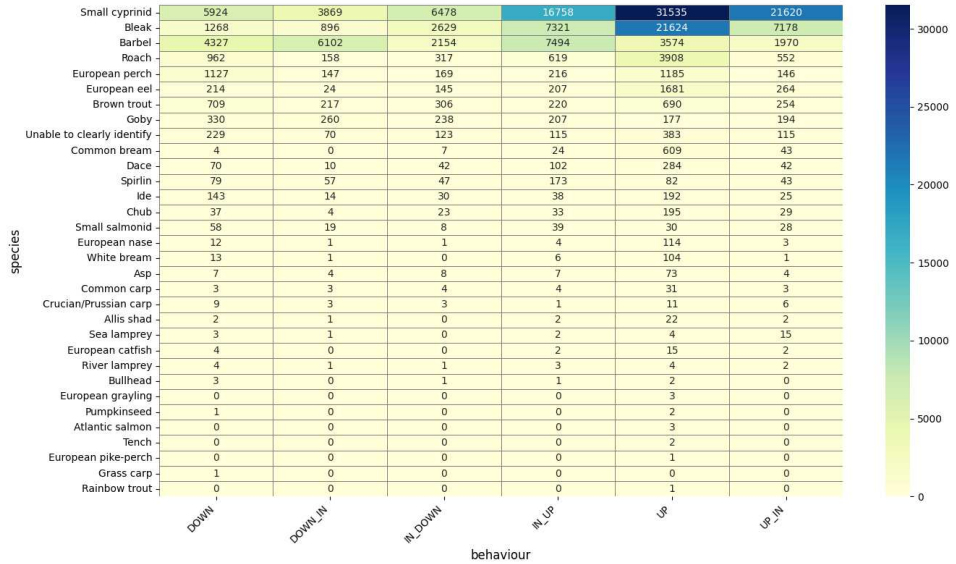
Proportional heatmaps should be interpreted with caution for rare species. When total observation counts are low, a small number of detections can produce extreme proportions (e.g. 100% in a single behaviour), which may not reflect consistent behavioural patterns. For this reason, raw count heatmaps are presented alongside proportional heatmaps to provide context on data availability.



(a) Site 1

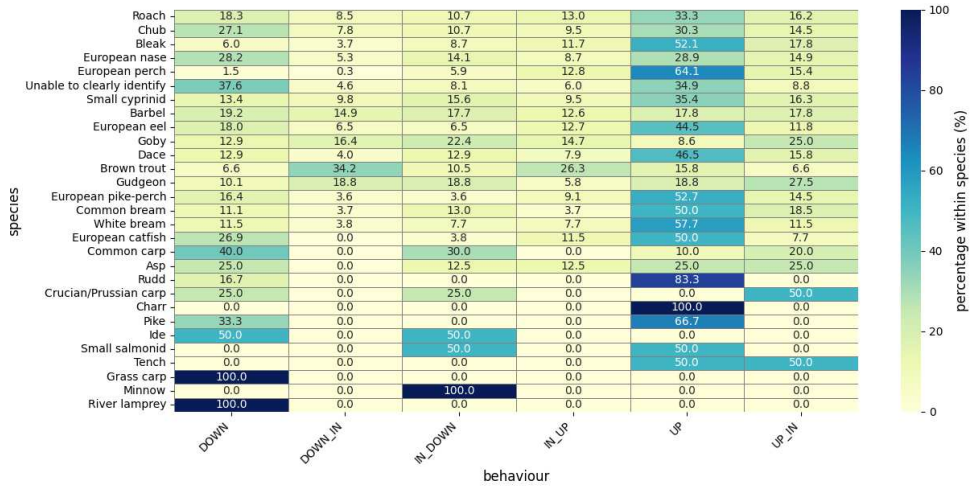


(b) Site 2

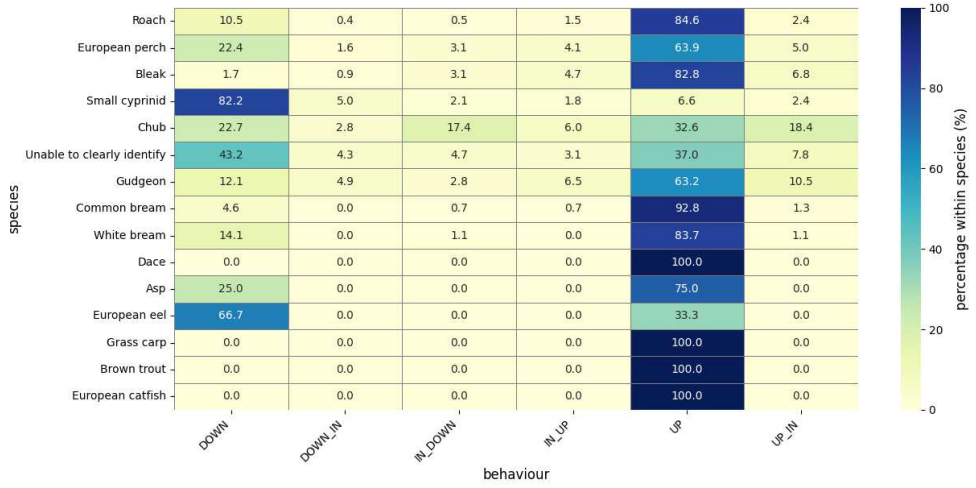


(c) Site 3

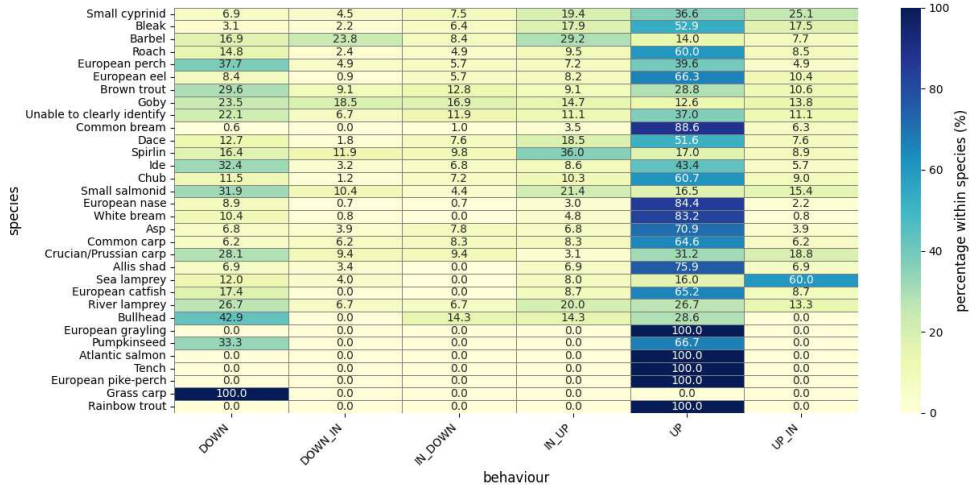
Fig. 3: Heatmaps showing the frequency of migration behaviours by species for each monitoring site. Darker cells indicate higher counts of a given species–behaviour combination.



(a) Site 1



(b) Site 2



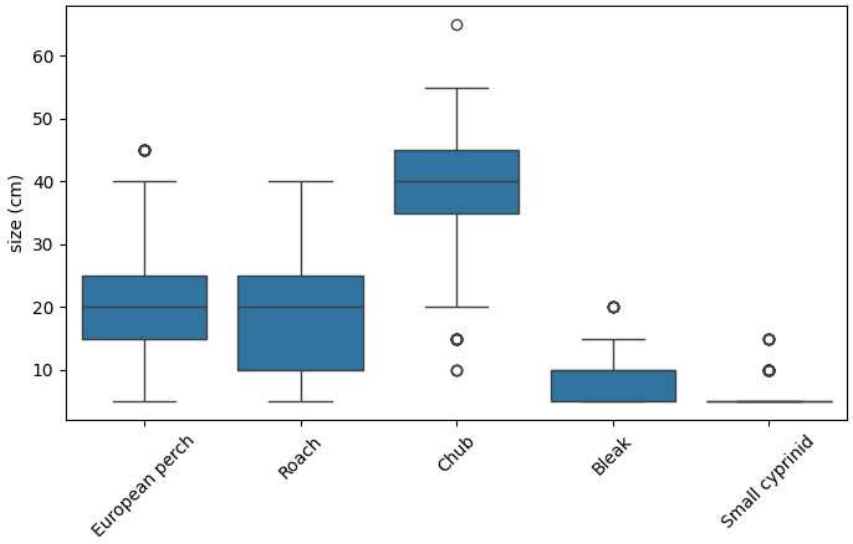
(c) Site 3

Fig. 4: Heatmaps showing species-normalised migration behaviour proportions for each site. Values represent the percentage distribution of behaviours within each species row-wise. Darker cells indicate higher proportions.

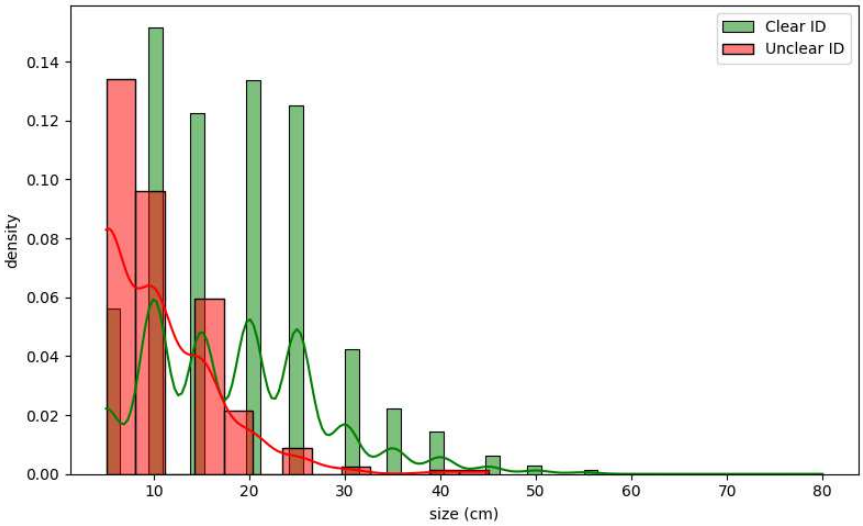
Taken together with the size–behaviour analysis (Section 3.1.1), these results show that size–behaviour patterns occur alongside differences in species composition. Downstream-oriented behaviours are more often associated with smaller-bodied individuals in these datasets. To support

the interpretation of species–size structure and identification uncertainty, Figure 5 illustrates size distributions by species and the relationship between size and identification clarity for Site 2.

These relationships are consistent with standard quality-control practices during annotation. For example, if a model assigns a species name to a fish outside of the typical size range for that species, this may indicate a potential classification error; for instance, a 5 cm fish is unlikely to be a Chub, while a 40 cm fish is unlikely to be classified as a Small Cyprinids. Behaviour can provide an additional layer of validation when considered jointly with species and size. For instance, certain combinations of species, size, and movement behaviour occur only rarely in the dataset and may therefore indicate potential annotation inconsistencies. These examples reflect general tendencies in the data and should be interpreted in the context of the monitored systems.



(a) Size (in cm) by species (top five, Site 2).



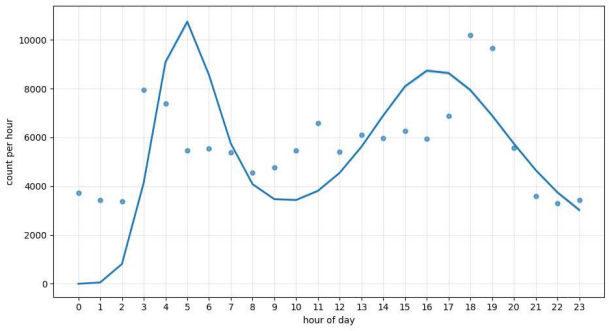
(b) Clear vs. unclear identifications (Site 2).

Fig. 5: Size patterns relevant to annotation interpretation, illustrating with Site 2. The most common species span distinct size ranges (top), and records labelled *Unable to clearly identify* are concentrated at the smallest sizes (bottom).

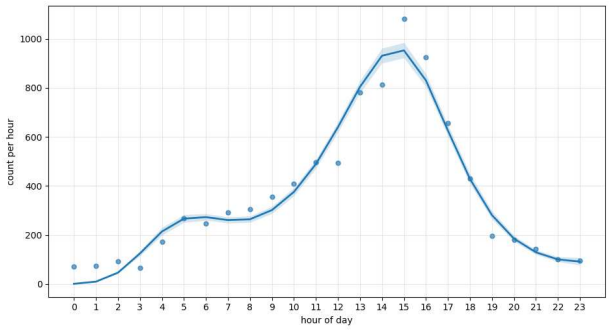
3.1.3 Temporal and behavioural structure of hourly fish passage

Hourly patterns of fish movement were analysed across all three monitoring sites to characterise diel structure and to identify consistent temporal signals relevant for later forecasting analyses. For each site, we examined (i) overall hourly activity, (ii) behaviour-specific diel patterns, (iii) size-structured activity, and (iv) species-specific activity using GAM-smoothed hourly counts together with observed counts (Figures 6–9). These patterns are defined with respect to clock time (hour of day) and do not account for seasonal variation in light conditions (e.g. sunrise and sunset times).

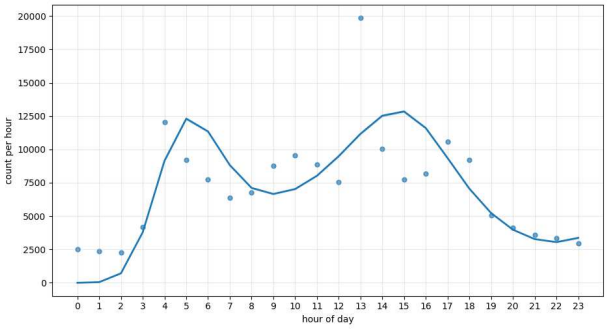
Overall diel activity across sites. Hourly fish counts varied systematically over the 24-hour cycle (Figure 6). The overall pattern is broadly *bimodal*, with activity lowest during hours 00:00–04:00, an initial peak around 04:00–06:00, followed by a decline and a second, broader peak around 14:00–17:00. This structure is most clearly expressed at Sites 1 and 3. At Site 2, the same pattern is present but less distinct, with activity dominated by a single peak around 14:00–15:00 and only a weaker increase before this peak. Overall, most observed passage is concentrated between approximately 04:00 and 18:00. Although overall activity levels vary substantially between sites, the timing of peaks is consistent across sites. Because counts are aggregated across the full monitoring period, the effective sample size contributing to each hour of day is large, resulting in relatively narrow confidence intervals around the mean diel curve. Site 2 generally showed a closer agreement between observed hourly counts and the fitted GAM curves than Sites 1 and 3, as also reflected by the lower NRMSE values (Table 3).



(a) Site 1



(b) Site 2



(c) Site 3

Fig. 6: Overall diel patterns of fish activity across the three monitoring sites. Points show observed hourly counts, while lines represent GAM-smoothed trends with 95% confidence intervals.

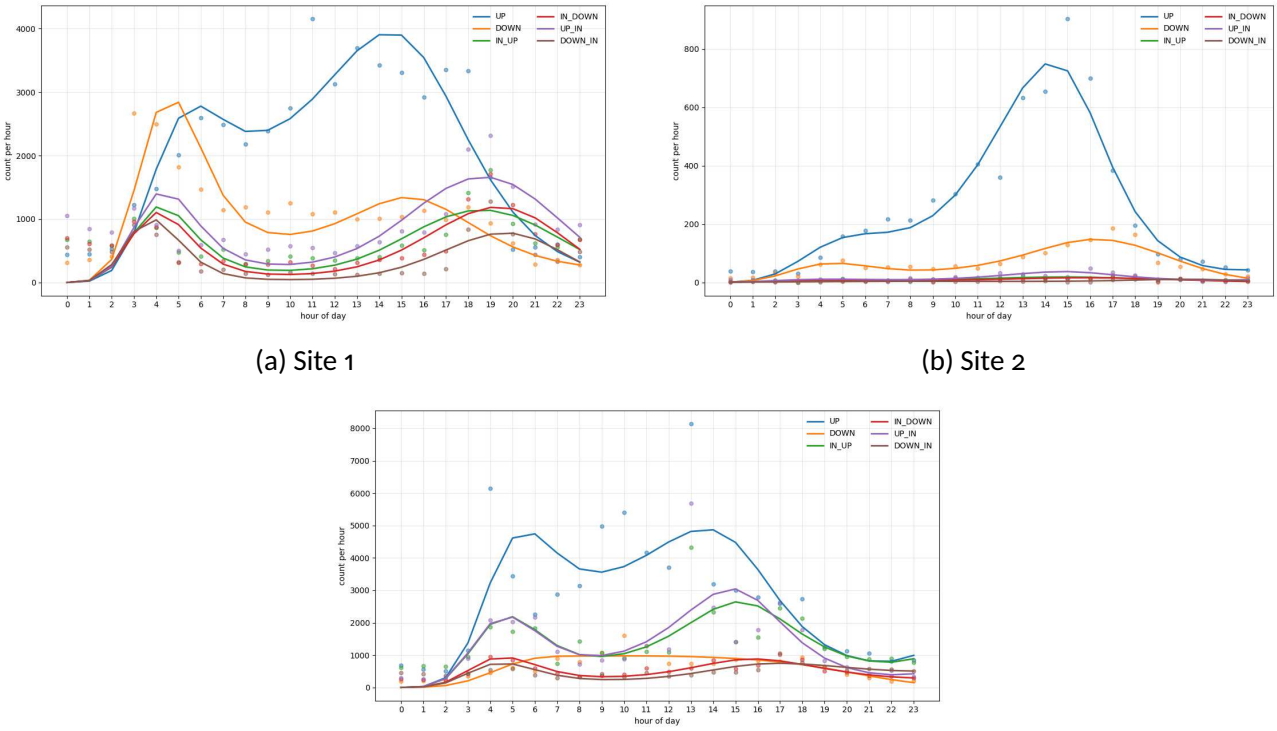
Table 3: Goodness-of-fit metrics for the GAM-smoothed overall diel curves at each monitoring site. RMSE denotes root mean squared error and NRMSE denotes RMSE normalised by the mean hourly count. Lower NRMSE values indicate closer agreement between observed hourly counts and the fitted GAM curve.

Site	RMSE	NRMSE
Site 1	2331.23	0.412
Site 2	62.36	0.171
Site 3	2830.87	0.393

Behaviour-specific diel patterns. Across all sites, the diel signal is dominated by complete upstream passages (`UP`) (Figure 7). This is in line with the monitoring setup, as the data are collected inside or near the exit of fishways designed to facilitate the ascent of fish across barriers, so a bias towards `UP` behaviour is expected. The `UP` curve closely tracks the total activity pattern, including the early-morning rise and the larger afternoon peak, indicating that most movement recorded by the cameras is full upstream passage.

Complete downstream passages (`DOWN`) are generally the second most frequent behaviour at Site 1, but they follow a different timing pattern than `UP`. In particular, `DOWN` shows a pronounced early-morning pulse (around 03:00–06:00) at Site 1, and a weaker but still visible morning increase at Site 2, followed by a broader afternoon/early-evening increase (roughly 14:00–18:00). This suggests that downstream movement contributes mainly to the morning peak and to the late-day shoulder, rather than driving the main afternoon maximum.

At Site 3, the pattern differs: although `UP` still dominates, the next most prominent signals come from partial behaviours (`UP_IN` and `IN_UP`), which rise strongly in the afternoon and approach the magnitude of `UP` for several hours. In contrast, `DOWN` remains comparatively smaller at this site. Overall, cross-site differences in diel shape are therefore influenced mainly by variation in the timing and strength of `UP` and `DOWN` (Sites 1–2), and by an additional strong contribution of partial movements at Site 3.



(a) Site 1

(b) Site 2

(c) Site 3

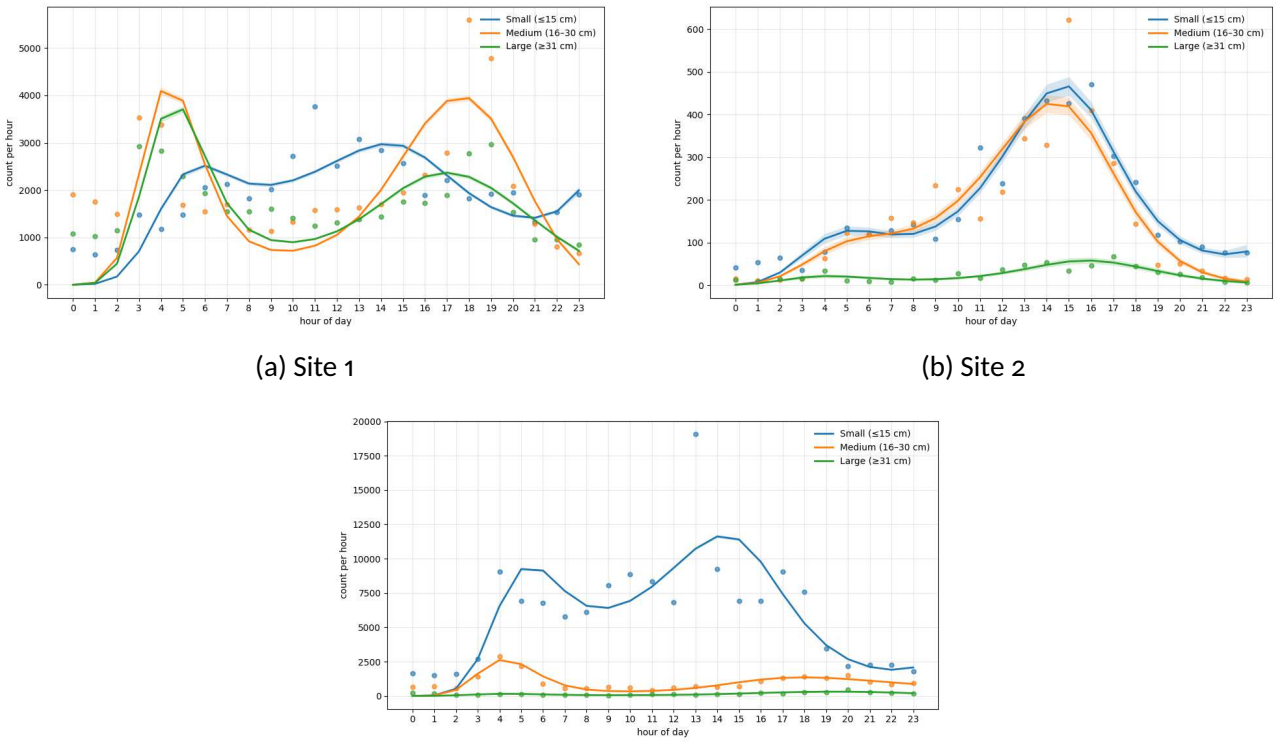
Fig. 7: Behaviour-specific diel patterns across sites. Points show observed hourly counts, while lines represent GAM-smoothed trends with 95% confidence intervals.

Size-structured diel activity. Grouping fish into three fixed size classes (Small $\leq$15 cm; Medium 16–30 cm; Large $\geq$31 cm) revealed consistent diel structure across sites, but with differences in the relative contribution of each size class (Figure 8).

Across sites, small and medium fish generally followed similar temporal patterns, with increasing activity during early hours of the day and more pronounced peaks later in the day, often strongest during the afternoon. The relative contribution of these two groups varied by site. At Sites 1 and 3, small fish dominated the overall counts and largely determined the shape of the aggregated diel curve, whereas at Site 2, medium-sized fish contributed more strongly and formed the dominant component of the diel signal.

Large fish were less abundant than the other size classes at all sites, but their relative contribution varied. At Sites 1 and 2, large fish exhibited clear diel variation with broader and lower-amplitude peaks compared with small and medium fish. In contrast, at Site 3, large fish remained rare throughout the 24-hour cycle, resulting in a comparatively weak and flattened diel pattern.

Overall, the dominant diel signal reflects the combined contribution of small and medium individuals, with the relative importance of each size class varying across sites, while large fish contribute more weakly to the temporal structure.



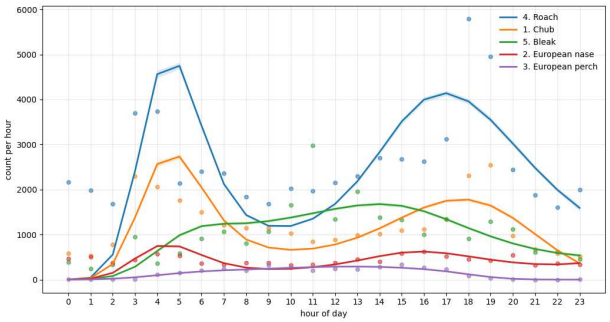
(a) Site 1

(b) Site 2

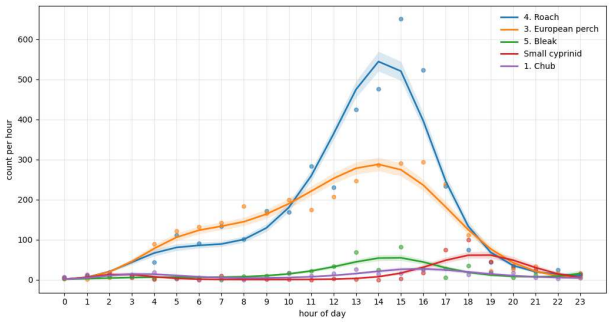
(c) Site 3

Fig. 8: Size-structured diel activity across sites. Points show observed hourly counts, while lines represent GAM-smoothed trends with 95% confidence intervals.

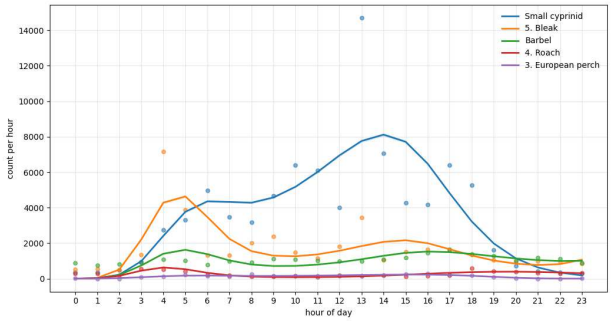
Species-specific diel patterns. Species-level analyses showed a consistent concentration of activity within a similar range of hours across sites, although the strength and shape of peaks varied among sites (Figure 9). At all sites, Cyprinids taxa, including Bleak, Small Cyprinids, and Roach, dominated observations and closely followed the overall and behaviour-driven diel patterns. Roach, which span both small and medium size classes, exhibited similar diel timing despite their broader size range. Some species, including Chub, Barbel, and European perch, occurred at lower frequencies or showed broader activity patterns and less sharply defined peaks, but largely within the same range of hours. Despite site-specific differences in species composition, cross-site temporal patterns largely share a diel structure, reflecting the activity of the dominant taxa (Cyprinids and percids) at each site.



(a) Site 1



(b) Site 2



(c) Site 3

Fig. 9: Species-specific diel activity for the most frequent species at each site.

Although diel patterns are defined at an hourly resolution, short-term forecasts depend strongly on the most recent observed hours. Because recording gaps frequently remove this immediate context, the model often must forecast the next 24 hours without knowing whether activity has recently increased, declined, or remained stable.

3.2 Forecasting performance

3.2.1 Example 24-h forecasts

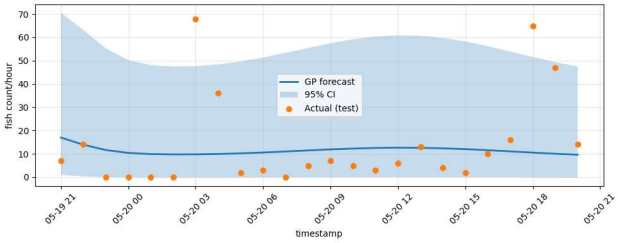
Figure 10 shows example 24-hour forecasts for Roach at all three sites using the densest near-end window at each site. In all sites, Roach counts in the selected windows are frequently sparse (many zeros punctuated by occasional non-zero hours). In these low-count conditions, the GP predictive mean tends to be smooth and conservative, while the predictive intervals widen to reflect uncertainty. This behaviour is expected: when the recent hourly series contains limited information about imminent spikes, the model prioritises the broader temporal structure learned from the historical record and expresses uncertainty rather than producing sharp peak forecasts (i.e. localising activity to the exact hour of occurrence).

Forecasts for the `Small` size group (Figure 11) generally show a stronger and more continuous signal than species-only Roach in the example windows, with more frequent non-zero observations. The predictive means more clearly track gradual within-day variations, but sharp bursts can still be underestimated, again with uncertainty widening around periods where the model cannot confidently localise short-lived peaks.

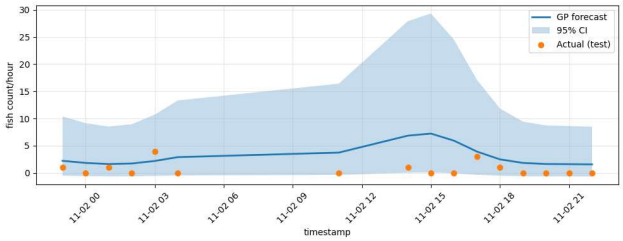
Finally, the *Roach*×`Small` target (Figure 12) illustrates the trade-off when targeting a more specific subgroup. Relative to `Small` alone, counts are typically lower and more intermittent, and the predictive mean becomes flatter while predictive intervals remain wide. In other words, conditioning on both species and size improves target specificity but can reduce predictability when the resulting time series is sparse.

Across all sites and targets, the GP produces smooth hourly trajectories with wider uncertainty bands when counts are low or highly variable, reflecting limited information for resolving

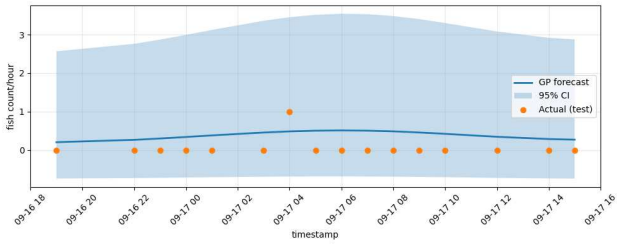
short-lived peaks within individual 24-hour test windows.



(a) Roach — Site 1

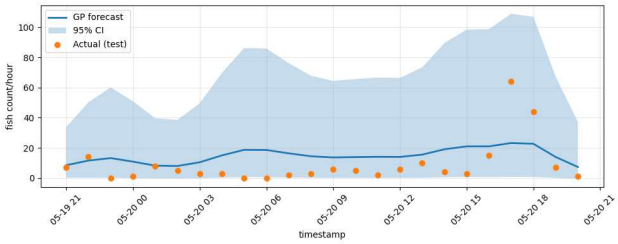


(b) Roach — Site 2

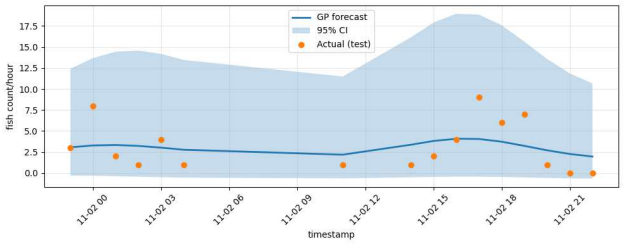


(c) Roach — Site 3

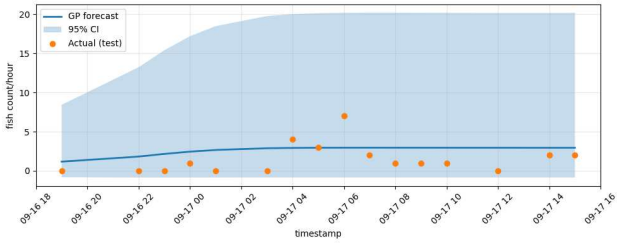
Fig. 10: Representative 24-hour GP forecasts for *Roach* across all sites, using the densest near-end 24-hour window at each site. Points show observed hourly counts in the test window; lines show GP predictive means; shaded bands indicate 95% predictive intervals.



(a) `Small` — Site 1

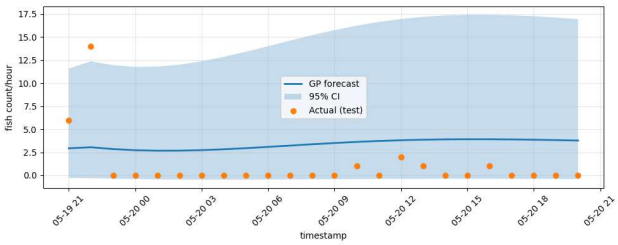


(b) `Small` — Site 2

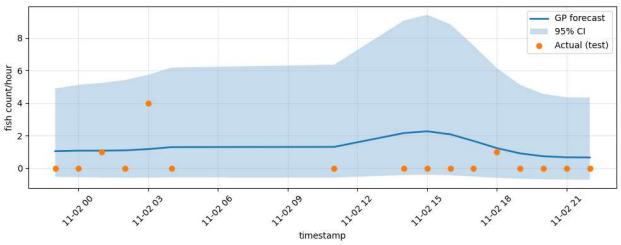


(c) `Small` — Site 3

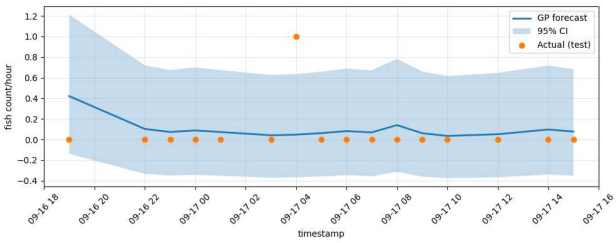
Fig. 11: Representative 24-hour GP forecasts for the `Small` size class (fish $\leq$15 cm) across sites (densest near-end 24-hour window per site).



(a) Roach×`Small` — Site 1



(b) Roach×`Small` — Site 2



(c) Roach×`Small` — Site 3

Fig. 12: Representative 24-hour GP forecasts for the combined *Roach*×`Small` target across sites (densest near-end 24-hour window per site).

The same model structure and evaluation approach are used throughout, but the effective test-set size differs because the test window contains only hours when video was available. This is visible in the plots as fewer points for sites with outages. The overall monitoring coverage per site is summarised in Table 1 and Table 4 reports the exact 24-hour windows used for the representative forecast plots and the corresponding observed-hour train/test sizes.

Table 4: Site-specific metadata for the representative 24-hour forecast windows shown in Figs. 10–12. Monitoring-period totals are reported in Table 1.

Site	Example test window (start → end)	Observed coverage (hrs/24)	n_{train} (obs. hrs)
Site 1	2021-05-19 20:00 → 2021-05-20 20:00	24/24 (100%)	2949
Site 2	2019-11-01 22:00 → 2019-11-02 22:00	16/24 (67%)	1021
Site 3	2023-09-16 18:00 → 2023-09-17 18:00	17/24 (71%)	4431

3.2.2 Rolling-window validation across sites

Table 5 summarises rolling-window validation for *Roach* across all sites (30 accepted windows per site). The median number of observed test hours differs among sites (Site 1: 18 [17–24]; Site 2: 18 [17–21]; Site 3: 23 [22–24]), reflecting differences in near-end recording continuity. This directly affects the effective test-set size because all metrics are computed only over observed hours.

Uncertainty calibration is consistently strong across sites: empirical coverage of the nominal 95% interval is high at all sites (median ≥ 0.88) and reaches 1.00 at Sites 1 and 3. This indicates that when the model is uncertain, for example during intermittent, spike-like passage, it typically expresses that uncertainty through wider predictive intervals rather than producing overconfident point forecasts.

Point accuracy summarised by MAE, varies markedly across sites, which is expected because MAE scales with the overall magnitude of hourly counts in the evaluation windows. Site 1 shows the largest and most variable MAE (12.92 [4.48–35.99]), consistent with some windows containing substantially higher Roach passage than the other sites. Site 2 shows moderate MAE (3.29 [2.25–4.47]) and a negative median TRE (-0.32 [-0.46 to 0.15]), indicating a tendency to

under-predict the 24-hour total in many windows, although the direction and magnitude of bias varies across windows. Site 3 has the lowest median MAE (0.88 fish/h), consistent with generally low counts in many windows; in these low-total regimes, TRE can nonetheless be highly variable because small denominators amplify relative differences.

Finally, correlations between observed and predicted hourly counts are low at all sites (medians near zero). This does *not* imply that the model fails to capture overall passage magnitude or uncertainty. Rather, it reflects the combination of (i) spike-dominated hourly dynamics with many zeros and occasional bursts, and (ii) a smooth GP prior (with periodic and RBF components) that tends to distribute probability mass over short-lived peaks unless the recent history provides clear evidence for their timing. For this reason, we treat correlation as a secondary diagnostic of within-window timing, while MAE, TRE, and interval coverage provide the main evidence for forecast performance in this setting.

Table 5: Rolling-window validation summary for *Roach*. Values are medians across accepted rolling 24-hour test windows near the end of each dataset, with interquartile ranges in brackets.

Site	Median n_{test} (obs. hrs)	MAE (median [IQR])	Timing corr. (median [IQR])	Total rel. error (median [IQR])	95% coverage (median [IQR])
Site 1	18 [17–24]	12.92 [4.48–35.99]	0.12 [-0.13–0.36]	0.31 [-0.22–0.63]	1.00 [0.92–1.00]
Site 2	18 [17–21]	3.29 [2.25–4.47]	-0.02 [-0.14–0.28]	-0.32 [-0.46–0.15]	0.88 [0.77–0.95]
Site 3	23 [22–24]	0.88 [0.58–1.93]	-0.12 [-0.25–0.03]	0.12 [-0.79–2.65]	1.00 [0.88–1.00]

4 Discussion

Using annotated underwater video from different monitoring sites as ground-truth observations, we characterised the statistical structure of hourly fish passage and explored how predictable hourly counts are from past observations. We examined diel timing, size–behaviour and species–behaviour structure, and a simple short-term forecasting baseline. Diel patterns were analysed using clock time (hour of day) rather than seasonally varying light phases (e.g. dawn, day, dusk, night). Because the dataset spans multiple seasons, the same clock hour may correspond to different ambient light conditions across the year, and the reported patterns should therefore be interpreted as average clock-time activity patterns rather than fixed responses to dawn or dusk conditions.

Across all sites, fish passage showed a clear diel pattern: most annotated passage occurred between early and late hours of the day, with an initial increase around 04:00–06:00 and a broader peak later in the day (around 14:00–17:00). This pattern was driven mainly by complete upstream movements and by numerically dominant Small Cyprinids taxa. Behaviour-, size-, and species-specific summaries mostly changed the strength and sharpness of peaks rather than shifting their timing, and the overall temporal organisation was similar despite strong differences among sites in composition, abundance, and recording continuity.

Size–behaviour structure was also consistent. Small fish appeared in every behaviour category and accounted for most annotations, while differences among behaviours were most evident in the upper tails of the size distributions. Downstream exit and partial upstream entry behaviours more often included larger individuals, whereas downstream entry-only behaviours were dominated by smaller fish. Species–behaviour results suggest that these patterns partly reflect which species are present at each site and their size ranges.

Taken together, this structure provides simple constraints that automated systems can use as reference patterns (e.g., diel timing and realistic species–size–behaviour combinations). The dominant pattern of upstream migration is partly influenced by the monitoring setup, as the data are collected within fishways designed to facilitate upstream passage across barriers [12, 45]; as a result, a dominance of `UP` behaviour is expected, while downstream-oriented movements may be underrepresented because such structures are not designed to be located or passed by downstream-migrating fish. In addition, the illuminated fishway environment may influence observed behaviour and reduce the detection of nocturnal activity, meaning that fish primarily active at night may be underrepresented [12, 46]. If that is the case, using these data to flag night-time detections could increase existing sampling bias when applied in machine learning

models. Annotation uncertainties (e.g. species misclassification or size estimation errors) may introduce further bias, which could be propagated or amplified when these data are used for modelling. Finally, fishway selectivity may influence which size classes are able to pass; in particular, there is evidence (e.g. at Site 2) that some installations may limit the upstream movement of smaller or juvenile fish [47]. As a result, the observed size distributions may be biased towards larger individuals and should not be assumed to represent typical migration patterns across all river systems.

We then used a Gaussian process model fit only to historical observed-hour counts to forecast the next 24 hours. Across sites and targets, the GP captured broad temporal structure while expressing uncertainty when recent observations provided little information about short-lived changes. When hourly counts were intermittent and contained many zero observations within recorded hours, together with occasional bursts of activity, the predictive mean often smoothed over isolated spikes, but these events were typically covered by the 95% predictive intervals. High empirical coverage across sites indicates that uncertainty was generally well calibrated. When information is limited, quantification of predictive uncertainty is an essential component of good forecasting practice [14]. Even in its simple form, this approach characterises expected temporal patterns and associated uncertainty when short-lived spikes cannot be predicted precisely. In practice, however, the direct use of such forecasts for day-to-day management in fishways is limited [48], since these structures are primarily designed to facilitate upstream passage rather than adaptive control. Instead, the main value of short-term forecasting in this context lies in validation and quality control. Expected diel patterns and predicted count ranges can be used to flag unusual observations or inconsistencies during annotation, potentially identifying errors in species classification, size estimation, or behaviour labels before they are added into the database.

An important limitation arises from how the data are collected. Because video recording is motion-triggered, zero-activity periods are underrepresented or absent, and the resulting time series is conditional on observed activity rather than continuous monitoring. As a consequence, the model is trained primarily on hours with fish detections and may therefore overestimate expected passage when interpreted as a continuous abundance signal. This reflects how the data are collected and should be kept in mind when interpreting the forecasts.

Differences in forecast sharpness and point error across sites were explained mainly by how many hours were actually observed within each 24-hour test window, rather than by total monitoring duration. In practice, each rolling evaluation used several hundred to several thousand observed training hours, so the results reflect limits of short-term predictability under sparse and irregular observation rather than insufficient training data. The target comparison also illustrates a clear trade-off: aggregated targets (e.g. the `Small` size class) produce more stable forecasts, whereas more specific targets (e.g. species–size combinations) become harder to predict as counts grow sparse. This trade-off is also useful for planning camera-based monitoring: if the goal is reliable day-ahead guidance for a particular (often rare) species or a narrow size class, then increasing effective coverage (more recorded hours per day, fewer gaps, or complementary sensors) will usually matter more than switching to a more complex model. Conversely, when coverage is limited or counts are low, the model can more reliably predict overall activity (e.g. total passage or broad size classes) than about individual species.

This study does not attempt to optimise spike prediction or to identify a model that “outperforms” others in capturing short-lived bursts. Instead, we use a simple, transparent forecasting baseline based only on past counts, in order to characterise the intrinsic limits of hourly predictability under intermittent observation, constrained in time by system dynamics and uncertainty [49, 50]. Here, transparency refers to the fact that the model structure is explicitly specified and interpretable: the periodic kernel captures diel cycles, while smooth components represent short- and longer-term temporal variation, and predictive uncertainty follows directly from the model assumptions. In contrast to highly parameterised approaches such as deep neural networks, whose internal representations may be difficult to interpret, the Gaussian process framework allows a clear link between assumptions, temporal structure, and resulting forecasts [51, 52]. This level of explainability is particularly important in monitoring contexts, where understanding why a forecast changes can be as important as the forecast itself.

In principle, additional covariates or models that explicitly represent regime shifts or event triggering, could improve localisation of short-lived peaks. However, the effectiveness of such approaches depends critically on the availability and continuity of both environmental data and recent observations. Our results therefore suggest that simply adding more historical observations of the same intermittent signal is unlikely to substantially improve short-term spike prediction. In contrast, improving recording continuity (i.e. reducing gaps in recent observations for e.g. camera systems record consistently during migration periods rather than relying heavily on motion-trigger thresholds that may exclude low-activity hours) or including relevant environmental covariates may provide additional information about event timing and are therefore more likely to improve short-term predictability than increasing model complexity alone. Many of these drivers are well established in fish migration studies; for example, spring migration is often closely linked to water temperature, downstream movements such as eel migration are triggered by discharge conditions, and juvenile dispersal can be influenced by oxygen availability and flow variability [53, 54]. In this context, automated forecasting systems should be evaluated not only by their ability to match exact spike timing, but also by the accuracy of 24-hour totals and by the calibration of their predictive uncertainty relative to known diel and behavioural structure.

Although this study focuses on hourly prediction, the results also inform what levels of temporal aggregation are likely to be reliable given typical monitoring data. Because fish passage often occurs in brief, clustered events, the timing of hourly spikes is influenced by the most recent observations. When forecasts rely only on past counts and recent hours are missing or uninformative, the exact timing of these events cannot be reliably predicted. At weekly or longer timescales, where individual spikes are averaged over, the consistent diel structure, size–behaviour relationships, and species composition observed across sites suggest that forecasts would be dominated more by structural patterns rather than short-term variability.

5 Conclusion

Fish passage across sites is characterised by consistent diel timing, size–behaviour relationships, and species composition, despite differences in monitoring conditions. These recurring patterns provide a practical reference for evaluating automated monitoring systems, as they define expected temporal structure and plausible species–size–behaviour combinations against which observations can be assessed.

Building on these patterns, the availability-aware forecasting baseline presented here helps to distinguish what can be learned from past observed-hour annotations from what remains intrinsically uncertain under intermittent observation. While broad temporal patterns can be captured, the exact timing of short-lived events depends strongly on recent observations and cannot be reliably predicted when data are sparse or missing.

At the same time, these patterns may be influenced by constraints of the monitoring system, including fishway design, illumination conditions, and annotation uncertainty, and should therefore be interpreted as system-specific observations. Short-term forecasting therefore provides a practical validation and quality-control tool, where expected temporal patterns and predicted count ranges can help to identify unusual observations or inconsistencies in annotation and automated outputs.

Future improvements in predictability are likely to depend less on additional historical observations and more on improved recording continuity and the inclusion of relevant environmental drivers of migration.

CRediT authorship contribution statement

Miracle Amadi: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, review & editing.

Jeffrey A. Tuhtan: Conceptualization, Investigation, Writing – review & editing.

David Nijssen: Investigation, Resources, Data curation, Writing – review & editing.

Cornelia Schütz: Investigation, Data curation, Writing – review & editing.

Data availability

The datasets analysed during the current study are not publicly available due to licensing and collaboration agreements, but may be available from the corresponding author upon reasonable request and with permission from the data owners.

Funding

This work was supported by the Finnish Cultural Foundation, South Karelia Regional Fund, through a postdoctoral research grant awarded to Miracle Amadi.

Jeffrey A. Tuhtan's contribution was supported in part by the Estonian Research Council grant PRG2198 and by the European Union and the Estonian Research Council through project TEM-TA141.

Acknowledgements

The authors thank the German Federal Institute of Hydrology (BfG) for providing the fishway monitoring datasets used in this study.

The authors also acknowledge the Environmental Sensing and Intelligence team at Tallinn University of Technology for support in data processing and annotation workflows.

The authors further thank Alexandra Kolosova, Paul Teesalu, Mihkel Treufeldt, Liina Toming, Ricardo Torres Cardona, and Sofia Orel for their contributions to the manual annotation of the fishway video data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Cooke, S. J. *et al*. Can the planetary health concept save freshwater biodiversity and ecosystems? *Lancet Planet. Health* **8,** e2–e3 (2024).
2. Dudgeon, D. *et al*. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews* **81,** 163–182 (2006).
3. Huang, M., Ding, L., Wang, J., Ding, C. & Tao, J. The impacts of climate change on fish growth: A summary of conducted studies and current knowledge. *Ecological Indicators* **121,** 106976 (2021).
4. De Santis, V., Jeppesen, E., Volta, P. & Korkmaz, M. *Impacts of human activities and climate change on freshwater fish—volume II* 2023.
5. Geist, J. Green or red: Challenges for fish and freshwater biodiversity conservation related to hydropower. *Aquatic Conservation* **31** (2021).
6. Deinet, S., Flint, R., Puleston, H., Baratech, A., *et al*. The Living Planet Index (LPI) for migratory freshwater fishes 2024 update: Technical report (2024).
7. Soom, J., Pattanaik, V., Leier, M. & Tuhtan, J. A. Environmentally adaptive fish or no-fish classification for river video fish counters using high-performance desktop and embedded hardware. *Ecological Informatics* **72,** 101817 (2022).
8. Castignolles, N., Cattoen, M. & Larinier, M. *Automatic system for monitoring fish passage at dams* in *Applications of digital image processing XVII* **2298** (1994), 419–429.
9. Michener, W. K. & Jones, M. B. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in ecology & evolution* **27,** 85–93 (2012).
10. Hampton, S. E. *et al*. Big data and the future of ecology. *Frontiers in Ecology and the Environment* **11,** 156–162 (2013).

11. Tuhtan, J. A., Nag, S. & Kruusmaa, M. Underwater bioinspired sensing: new opportunities to improve environmental monitoring. *IEEE Instrum. Meas. Mag.* **23,** 30–36 (2020).
12. Lucas, M. & Baras, E. *Migration of Freshwater Fishes* (John Wiley & Sons, 2008).
13. Nathan, R. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences* **105,** 19050–19051 (2008).
14. Dietze, M. C. *Ecological forecasting* (Princeton University Press, 2017).
15. Srivastava, G., Yoder, J. A., Park, J. & Kak, A. C. Using objective ground-truth labels created by multiple annotators for improved video classification: A comparative study. *Computer Vision and Image Understanding* **117,** 1384–1399 (2013).
16. Johnson, J. E., Patterson, D. A., Martins, E. G., Cooke, S. J. & Hinch, S. G. Quantitative methods for analysing cumulative effects on fish migration success: a review. *J. Fish Biol.* **81,** 600–631 (2012).
17. Shafait, F. *et al.* Fish identification from videos captured in uncontrolled underwater environments. *ICES Journal of Marine Science* **73,** 2737–2746 (2016).
18. Owen, A. & Emma, O. Real-time Fish Monitoring System Using Machine Learning (2024).
19. Atlas, W. I. *et al.* Wild salmon enumeration and monitoring using deep learning empowered detection and tracking. *Frontiers in Marine Science* **10,** 1200408 (2023).
20. Carpio, A. J., De Miguel, R., Oteros, J., Hillström, L. & Tortosa, F. S. Angling as a source of non-native freshwater fish: a European review. *Biological Invasions* **21,** 3233–3248 (2019).
21. Soom, J. *et al.* Open real-time, non-invasive fish detection and size estimation utilizing binocular camera system in a Portuguese river affected by hydropeaking. *Ecological Informatics* **90,** 103196 (2025).
22. Jansen, A. *et al.* Monitoring tropical freshwater fish with underwater videography and deep learning. *Marine and Freshwater Research* **75,** MF23166 (2024).
23. Ivanov, A. *et al. Freshwater Fish Detection for Video Tunnel in Fishways Under Real River Environments* Unpublished manuscript.
24. Beamish, F. W. H. in *Fish Physiology* (eds Hoar, W. S. & Randall, D. J.) 101–187 (Elsevier, New York, 1978).
25. Nathan, R. *et al.* A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* **105,** 19052–19059 (2008).
26. Elmes, A. *et al.* Accounting for training data error in machine learning applied to earth observations. *Remote Sensing* **12,** 1034 (2020).
27. Chandola, V., Banerjee, A. & Kumar, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)* **41,** 1–58 (2009).
28. Ruff, L. *et al.* A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE* **109,** 756–795 (2021).
29. Łopucki, R., Kiersztyn, A., Pitucha, G. & Kitowski, I. Handling missing data in ecological studies: Ignoring gaps in the dataset can distort the inference. *Ecological Modelling* **468,** 109964 (2022).
30. Carpenter, D., Deyle, E., Park, J., Saberski, E. & Sugihara, G. Repairing gaps in ecological time series. *Methods in Ecology and Evolution* **16,** 573–583 (2025).
31. Hastie, T. J. in *Statistical Models in S* 249–307 (Routledge, 2017).
32. Wood, S. N. *Generalized additive models: an introduction with R* (chapman and hall/CRC, 2017).
33. Pedersen, E. J., Miller, D. L., Simpson, G. L. & Ross, N. Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* **7,** e6876 (2019).
34. Seeger, M. Gaussian processes for machine learning. *International journal of neural systems* **14,** 69–106 (2004).

35. Schulz, E., Speekenbrink, M. & Krause, A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85,** 1–16 (2018).

36. Quinonero-Candela, J. & Rasmussen, C. E. A unifying view of sparse approximate Gaussian process regression. *Journal of machine learning research* **6,** 1939–1959 (2005).

37. Williams, C. K. & Rasmussen, C. E. *Gaussian processes for machine learning* **3** (MIT press Cambridge, MA, 2006).

38. Diggle, P. J., Tawn, J. A. & Moyeed, R. A. Model-based geostatistics. *Journal of the Royal Statistical Society Series C: Applied Statistics* **47,** 299–350 (1998).

39. Finney, D. On the distribution of a variate whose logarithm is normally distributed. *Supplement to the Journal of the Royal Statistical Society* **7,** 155–161 (1941).

40. Python Software Foundation. *Python Language Reference, version 3.12* 2024. https://www.python.org.

41. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12,** 2825–2830 (2011).

42. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: principles and practice* (OTexts, 2018).

43. Jones, P. E. *et al.* One size does not fit all: inter-and intraspecific variation in the swimming performance of contrasting freshwater fish. *Conservation Physiology* **8,** coaa126 (2020).

44. Rubio-Gracia, F., García-Berthou, E., Guasch, H., Zamora, L. & Vila-Gispert, A. Size-related effects and the influence of metabolic traits and morphology on swimming performance in fish. *Current Zoology* **66,** 493–503 (2020).

45. Bunt, C., Castro-Santos, T. & Haro, A. Performance of fish passage structures at upstream barriers to migration. *River research and applications* **28,** 457–478 (2012).

46. Helfman, G. S. Fish behaviour by day, night and twilight. *The behavioral ecology of teleost fishes*, 366–387 (1986).

47. Nijssen, D. *et al. Fischwanderung über eine Fischaufstiegsanlage - Erkenntnisse aus radiotelemetrischen Untersuchungen* in *Ergebnisse der Jahrestagung 2025 der Deutschen Gesellschaft für Limnologie (DGL) und der deutschen und österreichischen Sektion der Societas Internationalis Limnologiae (SIL)* (Frankfurt am Main, Germany, Sept. 2025).

48. Fu, C. *et al.* Making ecological indicators management ready: assessing the specificity, sensitivity, and threshold response of ecological indicators. *Ecological Indicators* **105,** 16–28 (2019).

49. Petchey, O. L. *et al.* The ecological forecast horizon, and examples of its uses and determinants. *Ecology letters* **18,** 597–611 (2015).

50. Makridakis, S., Spiliotis, E. & Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting* **36,** 54–74 (2020).

51. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* **1,** 206–215 (2019).

52. Molnar, C. *Interpretable machine learning* (Lulu. com, 2020).

53. Jungwirth, M., Schmutz, S., Weiss, S., *et al. Fish Migration and Fish Bypasses* (Fishing News Books, Oxford, 1998).

54. Bultel, E. *et al.* Migration behaviour of silver eels (Anguilla anguilla) in a large estuary of Western Europe inferred from acoustic telemetry. *Estuarine, Coastal and Shelf Science* **137,** 23–31 (2014).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial1.pdf